

Inputting method for telephone phonetic interactive system

Publication number: CN1531312

Publication date: 2004-09-22

Inventor: ZHANG JIYONG (CN); ZHU GUANG (CN); SUN WENYAN (CN)

Applicant: LEGEND BEIJING CO LTD (CN)

Classification:

- international: **H04L12/28; H04M1/23; H04M11/00; H04L12/28; H04M1/23; H04M11/00;** (IPC1-7): H04M11/00; H04L12/28; H04M1/23

- European:

Application number: CN20031019531 20030310

Priority number(s): CN20031019531 20030310

[Report a data error here](#)

Abstract of **CN1531312**

An input method for telephone voice interaction system relates to computer phone voice interaction system field, including following steps: the access device of telephone inputs the key signal to the command analysis control server, the server sends the voice signal into the speech recognition engine and the result of speech recognition is sent back to the command analysis server; the command analysis control server changes the key signal or speech recognition result into unified command; the voice application server feeds back the control signal to command analysis control server after receiving the unified command. This invention can utilize the speech or key to carry through input and can be used for mail, telephone number and stock, etc.

.....
Data supplied from the **esp@cenet** database - Worldwide

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷

H04M 11/00

H04M 1/23 H04L 12/28



[12] 发明专利申请公开说明书

[21] 申请号 03119531.8

[43] 公开日 2004 年 9 月 22 日

[11] 公开号 CN 1531312A

[22] 申请日 2003.3.10 [21] 申请号 03119531.8
[71] 申请人 联想（北京）有限公司
地址 100085 北京市海淀区上地信息产业基地创业路 6 号
[72] 发明人 张继勇 诸 光 孙文彦 任文捷
刘 武

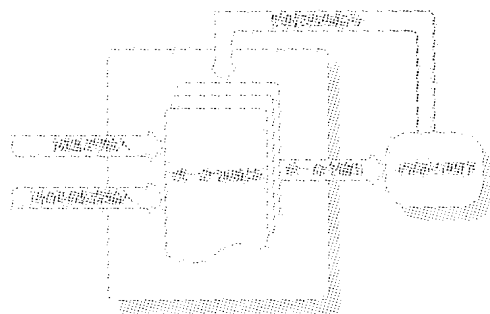
[74] 专利代理机构 北京集佳知识产权代理有限公司
代理人 王学强

权利要求书 2 页 说明书 14 页 附图 3 页

[54] 发明名称 电话语音交互系统的输入方法

[57] 摘要

本发明涉及计算机电话语音交互系统的技术领域，并具体公开了一种电话语音交互系统的输入方法。本发明的输入方法包括以下步骤：电话接入设备将接收到的按键信号输入至命令分析控制服务器，将接收到的语音信号输入至语音识别引擎，得到的语音识别结果输入至命令分析控制服务器；命令分析控制服务器将按键信号或语音识别结果转换为统一命令；语音应用服务器接收统一命令后反馈一控制信号至命令分析控制服务器。按照本发明，可实现自由利用语音或者按键进行输入，减少通话交互时间，操作使用准确方便。本发明的技术可应用于涉及邮件、电话号码、股票及其他各种信息服务领域。



1. 一种电话语音交互系统的输入方法，其特征在于，该方法包括以下步骤：

- 1) 将电话终端的按键信号或语音信号通过公用电话网输入至电话接入设备；
- 2) 若为按键信号，则电话接入设备将其输入至命令分析控制服务器；
- 3) 若为语音信号，则电话接入设备将其输入语音识别引擎，得到的语音识别结果输入至命令分析控制服务器；
- 4) 命令分析控制服务器将按键信号或语音识别结果转换为统一命令；
- 5) 语音应用服务器接收统一命令后反馈一控制信号至命令分析控制服务器。

2. 如权利要求 1 所述的电话语音交互系统的输入方法，其特征在于：步骤 4)中所述的转换通过命令映射转换单元实现，该命令映射转换单元由一组统一命令映射表组成。

3. 如权利要求 2 所述的电话语音交互系统的输入方法，其特征在于：步骤 1)之前，初始化当前系统状态和统一命令映射表。

4. 如权利要求 3 所述的电话语音交互系统的输入方法，其特征在于：语音应用服务器根据当前系统状态及统一命令得到下一系统状态。

5. 如权利要求 4 所述的电话语音交互系统的输入方法, 其特征在于:
根据下一系统状态得到步骤 5) 中所述的控制信号。
6. 如权利要求 5 所述的电话语音交互系统的输入方法, 其特征在于:
步骤 5) 之后, 刷新下一系统状态为当前系统状态。
7. 如权利要求 6 所述的电话语音交互系统的输入方法, 其特征在于:
命令分析控制服务器根据控制信号选定相应统一命令映射表。
8. 如权利要求 7 所述的电话语音交互系统的输入方法, 其特征在于:
步骤 5) 中所述的语音应用服务器根据当前系统状态和统一命令作出
响应, 输出语音信号或文本至命令分析控制服务器。
9. 如权利要求 8 所述的电话语音交互系统的输入方法, 其特征在于:
命令分析控制服务器输出语音信号至电话接入设备。
10. 如权利要求 8 所述的电话语音交互系统的输入方法, 其特征在于:
命令分析控制服务器输出文本经 TTS 语音合成引擎合成为语音信号
输入电话接入设备。

电话语音交互系统的输入方法

技术领域

本发明涉及计算机电话语音交互系统的技术领域，尤其涉及一种电话语音交互系统的输入方法。

背景技术

随着对各种自动化、智能化系统需求的不断增多，各种基于语音提示引导用户完成系统特定功能的电话语音交互系统日渐丰富，应用涉及邮件、电话号码、股票及其他各种信息服务领域。电话语音交互系统正逐渐成为一个新应用热点。

目前电话语音交互系统的输入有两种方法。

一种是传统的按键输入方法，即系统给出语音指令，提示用户应完成何种按键输入，同时接受用户的按键输入，然后执行相应的功能操作。这类系统一般采用 IVR (Interactive Voice Response) 式的树状结构，各个功能之间无法方便跳转，用户只能被动地听从系统的语音指令，使用起来很不方便，每次通话交互时间会很长。

另一种是语音输入方法。随着语音识别技术的不断成熟，语音输入作为一种重要的、新兴的输入方法得到了普遍的认可。语音输入方法具有快速定位的特点，能从大量的候选词中迅速确定输入的内容。比如对人名进行快速查找、在系统各个不同的功能之间进行快速跳转等。此外，这种输入方法操作方便自然，用户在使用时无需记忆复杂

的命令编号，只要直接说出自然语言的语音指令就能让系统执行相应的操作。正是由于以上的这些优点，语音输入方法正广泛应用于大量的系统中。

但是，语音输入具有一定程度的不确定性，也就是在某些情况下，语音识别结果可能出现错误，比如用户说“不是”，系统可能识别为“是”，从而导致误解。因此语音输入必须有一种确认机制：即根据语音识别结果的匹配程度，决定是否让用户对操作进行确认，可是采用了确认机制的语音输入方法又会使通话交互时间延长。

综上所述，现有的基于语音识别技术的电话语音交互系统中，片面强调了语音输入的优点，而没有注意到它的缺点。这些缺点如果不解决，则会让用户陷入频频出错的陷阱中，每次通话交互的时间非但不会减少，反而会延长，而且用户也会由于系统频频出错而放弃继续使用。

发明内容

本发明的主要目的是针对现有技术中存在的采用按键输入或语音输入的电话语音交互系统容易出错及通话交互时间长的缺陷而提供一种电话语音交互系统的输入方法。

为了达到所述目的，本发明电话语音交互系统的输入方法包括以下步骤：

- 1) 将电话终端的按键信号或语音信号通过公用电话网输入至电话接入设备；

- 2) 若为按键信号，则电话接入设备将其输入至命令分析控制服务器；
- 3) 若为语音信号，则电话接入设备将其输入语音识别引擎，得到的语音识别结果输入至命令分析控制服务器；
- 4) 命令分析控制服务器将按键信号或语音识别结果转换为统一命令；
- 5) 语音应用服务器接收统一命令后反馈一控制信号至命令分析控制服务器。

采用本发明电话语音交互系统的输入方法，可实现按键输入和语音输入有机结合的混合输入方法，一方面充分发挥了语音输入的优点，使得电话语音交互系统的输入非常人性化；另一方面把按键输入作为一个强有力的补充，有效克服了语音输入的不足。用户还可以根据自己的个人习惯、使用场合自由地利用语音或者按键进行输入，从而有效地解决了两种输入方法各自的缺点，并在很大程度上方便了用户的使用，不仅减少了通话交互时间，而且操作使用更加准确方便。

本发明随后将通过其具体实施例和附图加以说明。

附图说明

图 1 是采用混合输入方法的电话语音交互系统工作示意图。

图 2 是命令映射转换单元工作原理图。

图 3 是普通电话机的键盘布局图。

具体实施方式

本发明提出了一种电话语音交互系统的输入方法，此方法实现了按键输入和语音输入的有机统一，既确保了准确，又方便了使用。具体地说，就是本发明设计了一种混合输入的命令映射机制，使得语音信号与按键信号能按照一定的方式映射为统一命令，语音应用服务器上运行的语音应用程序只需处理这种统一命令，而无需关心是通过语音输入还是按键输入，从而方便了语音应用程序的设计。并在此基础上，通过对电话语音用户使用习惯的分析和研究，总结出电话语音交互系统的基本功能集合，而且针对这些功能设计了一套通用语音及按键输入方案，使得用户既可以方便的利用语音进行输入，也可以利用按键进行输入。下面分别对这两点进行详细介绍。

1. 混合输入的命令映射机制

请参照图 1 所示，采用混合输入方法的电话语音交互系统由电话接入设备、命令分析控制服务器、语音识别引擎、TTS 语音合成引擎（可选）以及语音应用服务器等部分组成。系统的具体输入方法包括以下步骤：

- 1) 电话终端的按键信号或语音信号通过公用电话网输入至电话接入设备；
- 2) 若为按键信号，则电话接入设备将其输入至命令分析控制服

务器；

- 3) 若为语音信号，则电话接入设备将其输入语音识别引擎，得到的语音识别结果输入至命令分析控制服务器；
- 4) 命令分析控制服务器将按键信号或语音识别结果转换为统一命令；此转换通过“命令映射转换单元”来实现。其原理如图 2 所示。

在这里，命令映射转换单元主要由一组“统一命令映射表”组成。按键信号的输入与语音识别结果的输入经过当前选定的统一命令映射表的处理后，映射成统一命令输送给语音应用服务器。

- 5) 语音应用服务器接收统一命令后反馈一控制信号至命令分析控制服务器。也就是语音应用服务器根据当前系统状态和统一命令得到下一系统状态，根据下一系统状态给出映射表控制信号，也就是统一命令映射表编号。命令分析控制服务器根据此统一命令映射表编号选定相应的统一命令映射表。
- 6) 刷新下一系统状态为当前系统状态。

下面解释一下系统状态和统一命令映射表。

应用于不同领域的电话语音交互系统可扩展成可执行特定功能的系统，针对不同系统的不同功能，我们应定义其所有可能的系统状态，如主状态、确认状态、退出状态以及执行具体功能的状态等。对于每个系统状态，我们又需要定义一个统一命令映射表。所以在系统运行前，在步骤 1) 之前存在一个初始化步骤，即根据系统自身需要

初始化系统状态和统一命令映射表。

统一命令映射表的格式如表 1 所示：

命令类型	命令内容	统一命令
KEY	0*	UM_HELP
SPEECH	帮助	UM_HELP
SPEECH	退出	UM_QUIT

表 1：统一命令映射表示例

其中“命令类型”栏包含按键（KEY）和语音（SPEECH）两种，“命令内容”栏存放具体的按键信号或语音识别结果，“统一命令”是一组事先定义好的常量。需要说明的是此统一命令映射表与系统状态是相关的，每个系统状态都有其对应的统一命令映射表。也就是说，在不同的系统状态下，相同的命令内容可以对应不同的统一命令。

命令映射单元将实际输入的按键信号或语音识别结果在“命令内容”栏中查找，如果匹配成功，则将其映射成相应的统一命令。当无匹配时，则将它们都映射为 UM_NOMATCH 命令。通过这种转换，语音应用程序就无需考虑具体的命令，而只需对统一命令进行处理即可。这极大地方便了语音应用程序的开发。

下面介绍一下语音应用服务器接收统一命令后的响应流程。首先语音应用程序根据当前系统状态和统一命令作出响应，输出语音信号或文本至命令分析控制服务器；命令分析控制服务器将语音信号直接输出至电话接入设备，将文本输出至 TTS（Text-To-Speech）语音合成引擎，经 TTS 语音合成引擎实时合成为相应的语音信号后输出至电

话接入设备；电话接入设备输出语音信号（包括事先录好的语音、合成的语音等）经公用电话网传送至电话终端。

2. 通用语音及按键输入方案

在上述的混合输入机制的基础上，本发明针对语音输入的不确定性，结合普通电话的键盘布局，设计出一套通用语音及按键输入方案，既有效、合理地分配了各个按键及它们的组合的功能，又具有易用、易记的特点。

从图 3 中我们可以看到，普通电话机的键盘布局非常简单，由 0～9 这 10 个数字键与两个功能键“*”、“#”组成。本发明的重点在于在不增加任何按键或者标识的前提下，将尽可能多的电话语音交互系统的功能命令以它们的组合来进行表示。

为了便于记忆和理解，首先我们对按键的两个功能键的作用分别进行定义：

键的作用：

1. 确定
2. 下一条
3. 表示顺序选择（与数字键组合）

* 键的作用：

1. 取消
2. 上一条
3. 表示功能选择（与数字键组合）

通过对电话语音交互系统的分析，我们总结出如下的一些所有电话语音交互系统都应具备的功能，并分别给出它们对应的语音输入和按键输入：

1. 操作确认功能（统一命令：UM_CONFIRM）

语音输入：确定

按键输入：#

说明：由于语音输入的不确定性，对上次的输入进行确认或者取消将是系统中使用很频繁的操作。语音输入选用“确定”和“取消”是因为它们的发音很容易区分，识别率高。同时，由于电话键盘只有两个功能按键，#键表示确定，*键表示取消也很容易让用户记住。

应用范围：操作确认的场合

2. 操作取消功能（统一命令：UM_CANCEL）

语音输入：取消

按键输入：*

说明：同上。

应用范围：操作确认的场合

3. 返回系统主界面功能（统一命令：UM_MAIN）

语音输入：重新开始

按键输入：**

说明：用户在使用电话语音系统的时候，由于系统的复杂性，有时会忘记自己正在执行的操作，或者由于错误的语音输入而执行了错误的操作，这时候重新开始的功能可为用户提供很大的方便。我们知道，*键表示取消功能，连续取消两次即回到系统的开始界面，很容易就能记住。

应用范围：所有场合

4. 退出功能（统一命令：UM_QUIT）

语音输入：退出系统

按键输入：***

说明：当用户需要退出电话语音系统时，除了直接挂机之外，还可以用命令来实现。*键表示取消功能，连续取消三次则表示退出系统。

应用范围：所有场合

5. 帮助功能（统一命令：UM_HELP）

语音输入：帮助

按键输入：0*

说明：在电话语音系统中，用户应该随时都能获得帮助信息。用户的语音命令为“帮助”，按键命令为“0*”。在这里，*键除表示取消外，还表示功能选择，0*表示系统的第0项功能，对应为帮助。

应用范围：所有场合

6. 系统功能选择（用户根据具体的功能定义相应的统一命令）

语音输入：相应的功能关键词（如：写信，转发，回信，等，它们是具体语音应用系统所需要提供的功能。）

按键输入：%*（注：%表示任意的数字键及其组合，下同）

说明：%键表示功能选择，数字对应的功能编号。由于具体的功能和具体的应用有关，因此用户需要记住各个功能所对应的编号，或者通过帮助来即时了解系统的功能编号。

应用范围：用户需要选择功能场合

7. 绝对顺序选择（统一命令：UM_ABSORDER_%）

语音输入：第%条（，封，人，天...）

按键输入：%#

说明：%除了表示确定外，其直观的意思是序号，因此用数字加上井号键表示绝对顺序选择，用户很容易记住。

应用范围：系统提示用户进行选择的场合

8. 选择上一条记录（统一命令：UM_PREVIOUS）

语音输入：上一条（，封，人，天...）；

按键输入：%*

说明：%键与#键在键盘上一左一右，并且%键在左边，类似向

前翻页，#键在右边，类似于向后翻页。两者结合起来很容易记忆

应用范围：系统提示用户进行选择的场合

9. 选择下一条记录（统一命令：UM_NEXT）

语音输入：下一条（，封，人，天...）；

按键输入：#

说明：同上

应用范围：系统提示用户进行选择的场合

总结来说，上述的各项功能及其对应的通用语音及按键输入如表2所示。

操作	语音输入	按键输入	应用范围
操作确认	确定	#	操作确认场合
操作取消	取消	*	
帮助	帮助	0*	所有场合
回系统主界面	重新开始	**	
退出系统	退出系统	***	
系统功能选择	各功能对应的关键词	%*	用户需要选择功能的场合
绝对顺序选择	第%条（，封，人，天...）	%#	系统提示用户进行选择的场合
选择上一条记录	上一条（，封，人，天...）	*	

选择下一条记录	下一条(, 封, 人, 天...)	#	合
---------	-------------------	---	---

表 2：通用语音及按键输入方案

下面我们以“电话语音邮件系统”为例来详细阐述本发明的优选实施例。

首先，我们定义此系统需要实现的具体功能。在“电话语音邮件系统”中，我们需要实现如下的功能：

需要实现的功能	对应的按键操作	统一命令
重播	1*	UM_REPEAT
写信	2*	UM_WRITEMAIL
转发	3*	UM_TRANSMITMAIL
回信	4*	UM_REPLYMAIL
修改密码	5*	UM_CHANGEPWD

表 3：电话语音邮件系统需要实现的功能列表

其次，我们还需要定义系统所有可能的系统状态。“电话语音邮件系统”可能处在的系统状态有：登录状态、主状态、写邮件状态、转发状态、确认状态、退出状态等。对于每个系统状态，我们需要定义一个统一命令映射表。例如系统在确认状态时的统一命令映射表内容如下：

命令类型	命令内容	统一命令
KEY	#	UM_CONFIRM
KEY	*	UM_CANCEL
KEY	0*	UM_HELP

KEY	**	UM_MAIN
KEY	***	UM_QUIT
SPEECH	确定	UM_CONFIRM
SPEECH	取消	UM_CANCEL
SPEECH	帮助	UM_HELP
SPEECH	重新开始	UM_MAIN
SPEECH	退出系统	UM_QUIT

表 4：系统确认状态时的统一命令映射表

按键信号或语音识别结果输入经统一命令映射表处理后输出统一命令至语音应用程序，语音应用程序根据当前系统状态和统一命令确定下一系统状态，根据下一系统状态反馈映射表控制信号，根据控制信号选定下次输入所应用的统一命令映射表，同时刷新下一系统状态为当前系统状态。

此外，电话语音邮件系统还需要在实现上述定义的各项基本功能的基础上，扩充一些新的功能，比如帐号输入，语音录音等，对于这些与应用相关的特定输入方式，需要特殊处理，无法采用统一命令。

最后需要说明的是，实际的电话语音应用程序的硬件部署不必完全参照图 1，可以根据实际情况将一些模块合并在同一台硬件设备上。例如语音识别引擎和 TTS 语音合成引擎可以合并到一台服务器上运行，命令分析控制服务器所执行的功能可以放在语音应用服务器上运行。

应用本发明能为电话语音交互系统开发人员和用户都带来好处。

对于系统开发人员来说，只要实现了上述的混合输入方法，语音应用程序就能有非常友好的用户使用界面；对于用户来说，只要掌握了上述的通用语音及按键输入方案，就几乎能立刻方便使用采用此输入方案的电话语音交互系统。

本发明不限于以上描述，只要不违背本发明的思想，在本发明的精神和原则之内所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

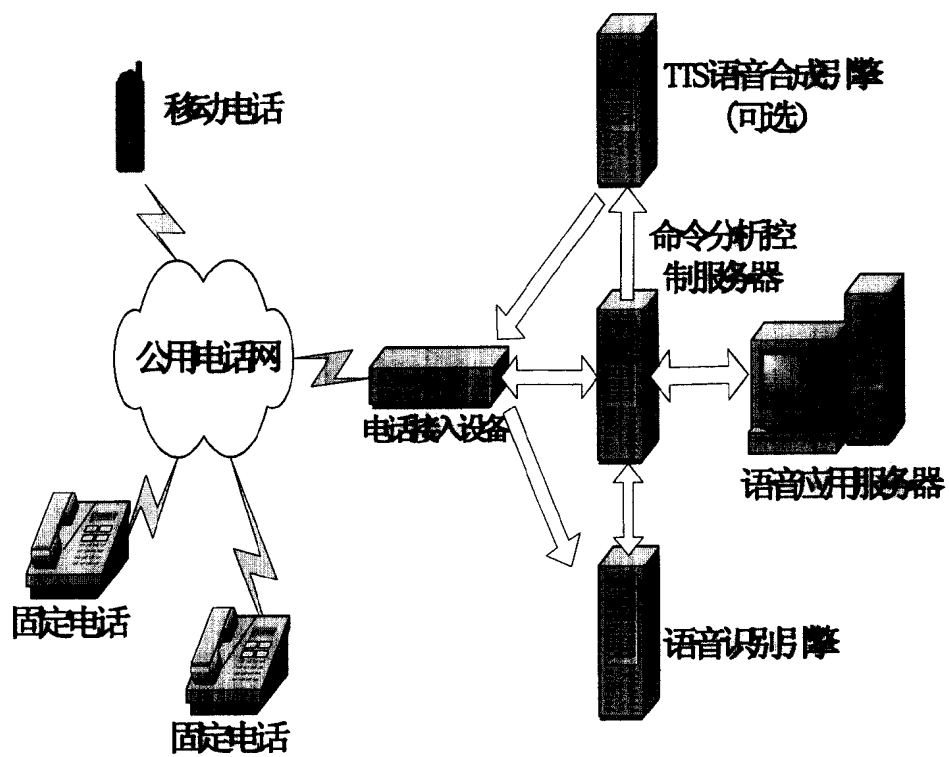


图 1

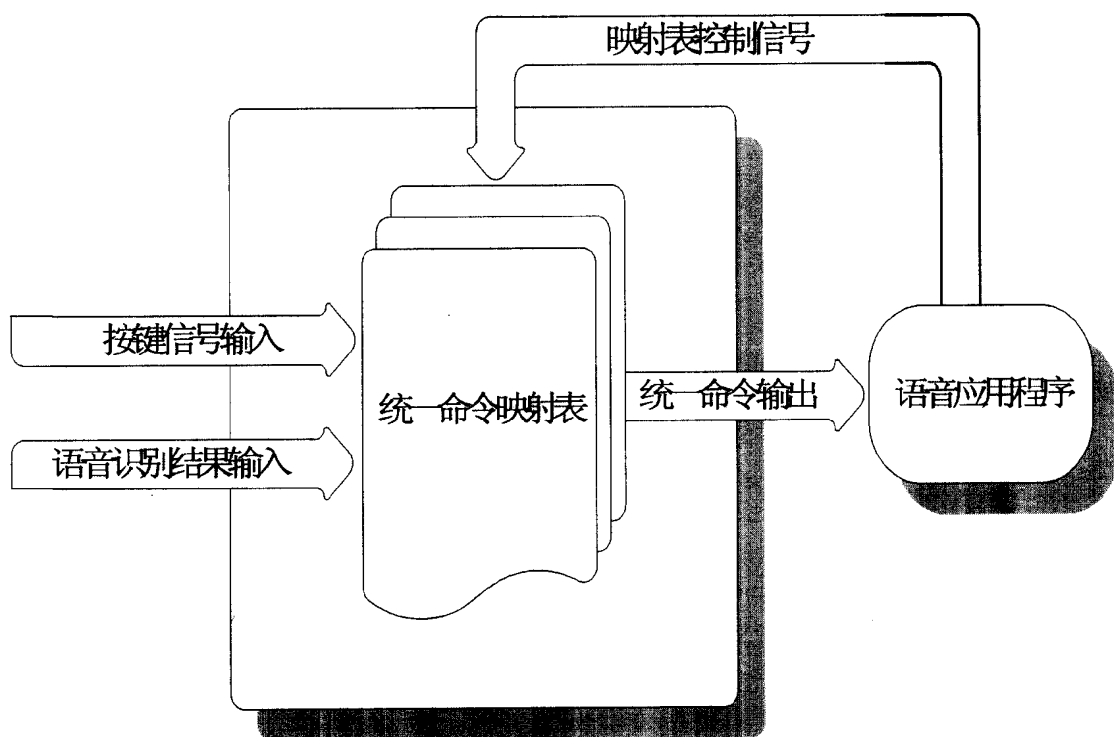


图 2

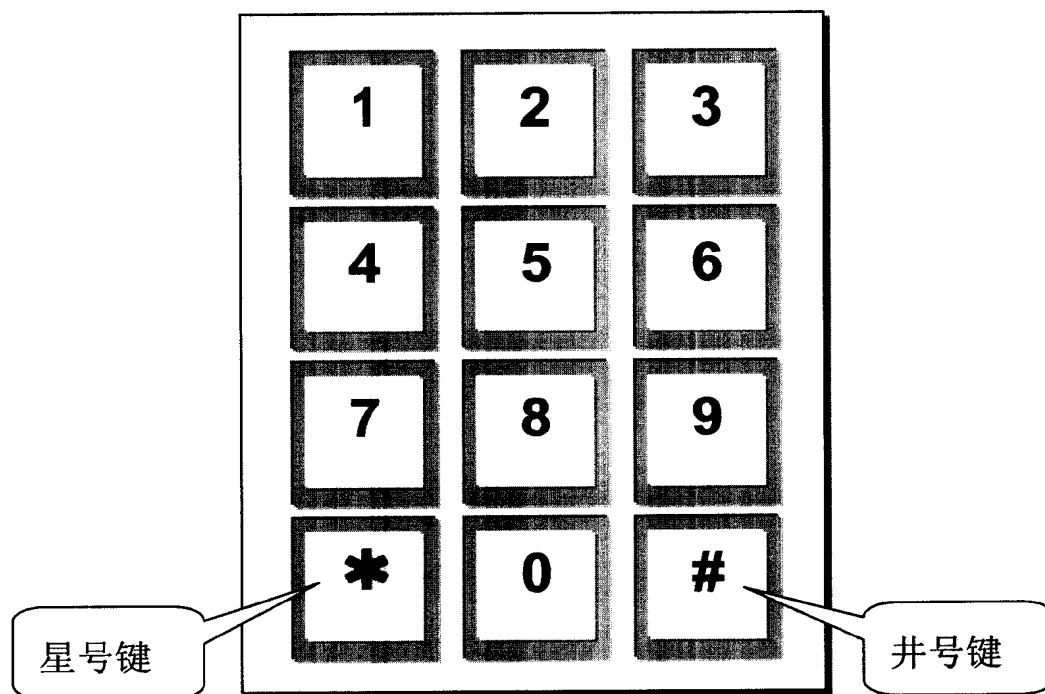


图 3